In search of Leonardo: Computer-based facial image analysis of Renaissance artworks for identifying Leonardo as subject

Christopher W. Tyler,[a] William A. P. Smith[b] and David G. Stork[c]

[a]Smith-Kettlewell Eye Research Institute, 2318 Fillmore Street, San Francisco, CA 94115 USA
[b]Department of Computer Science, The University of York, Heslington, YO10 5GH, York, UK
[c]Rambus Labs, 1050 Enterprise Way Suite 700, Sunnyvale, CA 94089 USA

ABSTRACT

One of the enduring mysteries in the history of the Renaissance is the adult appearance of the archetypical "Renaissance Man," Leonardo da Vinci. His only acknowledged self-portrait is from an advanced age, and various candidate images of younger men are difficult to assess given the absence of documentary evidence. One clue about Leonardo's appearance comes from the remark of the contemporary historian, Vasari, that the sculpture of *David* by Leonardo's master, Andrea del Verrocchio, was based on the appearance of Leonardo when he was an apprentice. Taking a cue from this statement, we suggest that the more mature sculpture of *St. Thomas*, also by Verrocchio, might also have been a portrait of Leonardo. We tested the possibility Leonardo was the subject for Verrocchio's sculpture by a novel computational technique for the comparison of three-dimensional facial configurations. Based on quantitative measures of similarities, we also assess whether another pair of candidate two-dimensional images are plausibly attributable as being portraits of Leonardo as a young adult. Our results are consistent with the claim Leonardo is indeed the subject in these works, but we need comparisons with images in a larger corpora of candidate artworks before our results achieve statistical significance.

1. INTRODUCTION

Scholarly opinion varies as to the number of portraits of Leonardo that exist. The red chalk portrait of an old man in the Milan collection of Leonardo's papers (Fig. 1a) is the only one acknowledged by Kenneth Clark, a foremost authority on Renaissance art and Leonardo in particular. Even this identification had been questioned, because the subject appears older than his early sixties, which are the latest that Leonardo could have been served as model in Italy, since he left for France at age 63. Perhaps, then, this is a portrait of his father Ser Piero da Vinci or his uncle Francesco. These doubts may be resolved by the fact that Leonardo was described by contemporaries as looking ten years older than his actual age, consistent with the apparent age of the sitter in the portrait. However, Leonardo's appearance when younger is generally regarded as obscure.

Even though only one authoritative portrait is recognized, there is an increasing set of portraits purporting to be of Leonardo that are gradually gaining acceptance among Renaissance art scholars. Restricting consideration to contemporary portraits, just three are known (Fig. 1): one by Leonardo's loyal pupil, Count Francisco Melzi, one by Raphael (who knew the artist from Leonardo's visits to his father Giovanni Sanzio in Urbino), and a newly discovered work by French stained-glass artist Guillaume de Pierre di Marcillat (who had relocated to Arezzo many years earlier, where in fact he was Giorgio Vasari's master).

Leonardo's appearance in middle age and when young is generally regarded as unknown, although he was described by Vasari[1] as of "outstanding physical beauty." Perhaps the most compelling potential portrait is the statue of David by Verrocchio, stated by Vasari to have been modeled on the youthful Leonardo.[1] This claim is not often taken too seriously by art historians (such as the above-mentioned Clark and Vezzosi), but is given plausibility by the fact that Leonardo was well known as Verrocchio's favorite pupil, and was just the right age for that role.

Taking a cue from this statement, we wondered whether the more mature sculpture of *St. Thomas*, constructed when Leonardo was Verrocchio's main assistant, could also have been a portrait of Leonardo (with the accompanying *Christ* as a self-portrait of Verrocchio). The intended interpretation could be that Verrocchio was representing himself as effectively baptizing Leonardo as the avatar of the next generation of artists. Finally,

(a) (b) (c) (d)

Figure 1. The most securely established contemporary portraits of Leonardo: a) *Self-portrait*, b) *Plato/Leonardo*, by Raphael (1510), c) *Leonardo*, by Guillaume de Pierre di Marcillat (1520), and d) *Leonardo*, by Francesco Melzi (ca. 1510).

there are at least two further portraits which, it is argued, may have Leonardo as their subject. In a video essay, Woldhek[2] analyzes 120 depictions of faces in Leonardo's work, and uses an iconographic criterion to exclude most of them, resulting in just three faces that he considers to be plausible self-portraits by Leonardo: the aged self-portrait drawing of Fig. 1a, the famous drawing of the Vitruvian man (Fig. 2d) and the *Portrait of a Musician* (Fig. 2c). This final portrait is puzzling, for a number of reasons.

Originally listed as *Portrait of Ludovico il Moro* (Ludovico Sforza, Duke of Milan, who was born in the same year as Leonardo), and as painted by Bernardo Luini (one of Leonardo's closest adherents), the painting has often been attributed as a portrait of Leonardo. A cleaning in 1905 revealed that the sitter is holding a piece of sheet music on which can be seen the letters "CANT...ANG..." (most likely meaning "songs of angels"), which would be inconsistent with the identification of the subject as Duke Ludovico, but would fit with the angelic persona attributed to the young Leonardo. Moreover, Leonardo had a great reputation for being able to play any stringed instrument at first sight. Thus, there seems to be strong evidence supporting the identification of the musician portrait as being of Leonardo, regardless of whether it was painted by Leonardo himself or by a contemporary.

In this paper we use a statistical method for three-dimensional face shape estimation to quantitatively evaluate the similarity between the faces in the four candidate portraits described above. We show that the face shape estimates from the four portraits cluster well, with smaller intra-class distances than inter-class distances to a set of ten distractor faces. In Sect. 2 we describe our computer vision method for comparing shapes, such as faces, based on morphable models and in Sect. 3 we show our results on candidate Leonardo portraits. We conclude in Sect. 4 and describe future directions.

2. ESTIMATING FACIAL SIMILARITY

Reconstructing three-dimensional face shapes from a single view image is a well-studied problem. In general, single-image shape estimation is an ill-posed problem even when strong assumptions are made about scene geometry, reflectance and camera properties.[3,4] In the case of paintings, this process is further confounded by potential variations introduced by the artist in their rendition of perspective, shading and illumination.[5] For these reasons, in this paper we neglect photometric cues and reconstruct three-dimensional face shapes using the projected position of fiducial points only. This approach has been shown to perform well on synthetic data.[6] We employ a linear statistical model of three-dimensional face shape (morphable model[7]), learned from a representative sample of human faces. The model is used to constrain the face shape recovery process by transforming it to one of parameter estimation. In contrast to Aldrian and Smith,[6] we use a weak perspective camera model. We find that this is more robust than an affine camera in handling perspective distortions introduced by the artist. Hence, face shape is linear in the morphable model parameters and alternating least squares is used to iteratively solve for pose and shape.

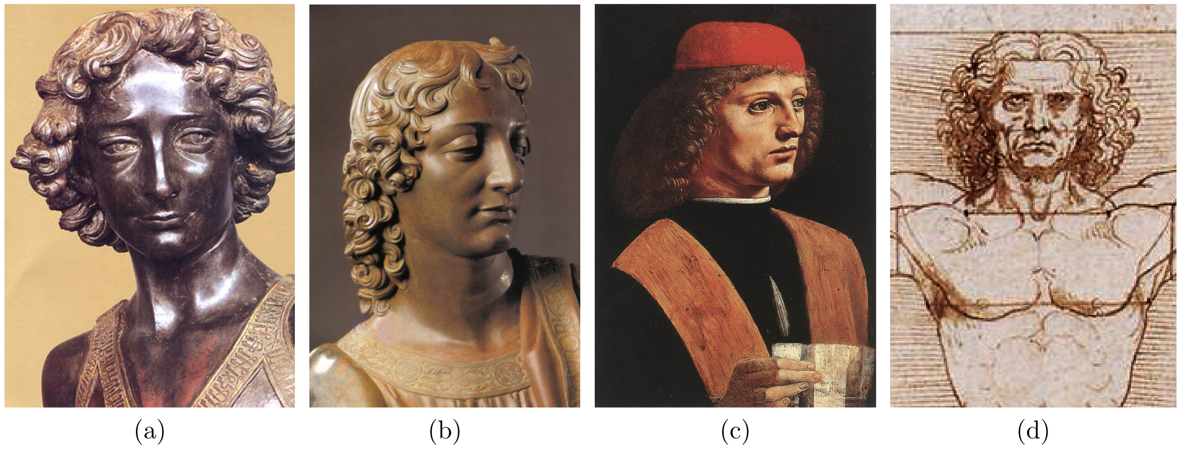

Figure 2. Candidate portraits of a young or middle-aged Leonardo: (a) *David* (Verrocchio), (b) *St. Thomas* (Verrocchio), (c) *Portrait of a musician* (Leonardo), (d) *Vitruvian man* (Leonardo).

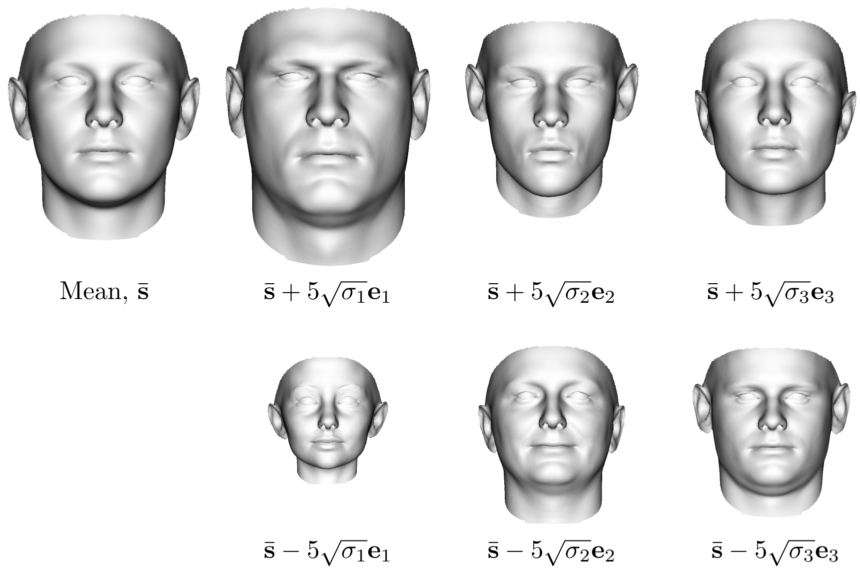

Figure 3. A 3D morphable model of human faces.[9] The top left panel shows the mean face surface. The remainder show the mean face deformed by ± 5 standard deviations along the first three principal modes of variation, $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$.

For robustness, we learn an empirical model of generalization error over the surface of the face. Such a model describes the variance between a true face shape and its best model fit. This approach allows for feature-sensitive weighting of data-closeness errors and allows us to regularize the optimization problem without manual parameter tuning.[8]

2.1 3D Morphable Model

A 3D Morphable Model (3DMM) is a parametric deformable shape model. The parameter space of the model is learnt via a statistical analysis of a sample of training data. They are particularly well suited to describing the class of human face surfaces may also be represented by linear statistical models.

In order to construct a 3DMM, meshes in the training data are set into dense correspondence. Hence, each training sample can be represented by a vector of length $3n$ formed by concatenating the x, y and z coordinates of the n vertices in the mesh: $\mathbf{s}_i = [x_1\ y_1\ z_1\ \ldots\ x_n\ y_n\ z_n]^t$. The ith vertex in each mesh corresponds to the

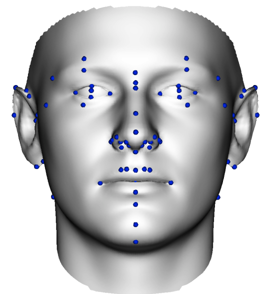

Figure 4. The 70 Farkas feature points shown on the mean face.

same location on the face. Principle components analysis is used to find the principal axes of variation within this high-dimensional space of face shapes. The resulting model allows any face shape to be approximated as a linear combination of the mean face and the modes of variation:

$$\mathbf{s} = \bar{\mathbf{s}} + \sum_{i=1}^{k} b_i \mathbf{e}_i = \mathbf{P}\mathbf{b} + \bar{\mathbf{s}}, \tag{1}$$

where $\bar{\mathbf{s}}$ is the mean face mesh and $\mathbf{e}_i$ is the ith principal component. The matrix $\mathbf{P} = [\mathbf{e}_1|\dots|\mathbf{e}_k]$ is formed by stacking the principal components. The variance along the ith principal component is given by the corresponding eigenvalue, σ_i. We refer to the ith vertex in the mesh given by the parameter vector $\mathbf{b}$ as $\mathbf{v}_i(\mathbf{b}) \in \mathbb{R}^3$. Figure 3 shows a representation of the first three modes of variation of a morphable model.

2.2 Model Fitting

We use a set of feature points proposed in the anthropometrics literature (Fig. 4).[10] This set of canonical points on the face surface were chosen because of their saliency in describing face morphology. For this reason they prove a good choice for reconstruction of faces from sparse data. For fitting to faces in painted works we use the visible subset of these points. For statues we use multiple images, using the visible subset in each image. In both cases, we augment the set of points via an iterative manual process. This allows errors in the reconstruction to be corrected by adding additional feature points where needed (typically along the occluding boundary). For a particular subject, the set of feature points in the kth image is represented by the set of feature point indices $\mathcal{F}_k$. The ith observed two-dimensional image coordinates of a feature point in the kth image is denoted $\mathbf{f}_{i,k} \in \mathbb{R}^2$.

We assume a weak perspective model. The two-dimensional position of the ith model vertex is determined by the shape parameters, $\mathbf{b}$, and the camera projection parameters, $\Theta = (\mathbf{R}, \mathbf{t}, s) \in SO(3) \times \mathbb{R}^2 \times \mathbb{R}$:

$$\mathbf{w}_i(\Theta, \mathbf{b}) = s\mathbf{P}\mathbf{R}\mathbf{v}_i(\mathbf{b}) + \mathbf{t}, \tag{2}$$

where

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{3}$$

is an orthographic projection matrix. Hence, the total error between a predicted and observed set of feature point positions for the kth image is given by:

$$\varepsilon(\Theta, \mathbf{b}) = \sum_{i \in \mathcal{F}_k} \left\| \mathbf{V}_i \left(\mathbf{w}_i(\Theta, \mathbf{b}) - \mathbf{f}_{i,k} \right) \right\|^2, \tag{4}$$

where the matrix $\mathbf{V}_i$ scales the error according to the predicted two-dimensional variance of the ith feature point:

$$\mathbf{V}_i = \begin{bmatrix} \frac{1}{\sigma^2_{2D,i,x}} & 0 \\ 0 & \frac{1}{\sigma^2_{2D,i,y}} \end{bmatrix}. \tag{5}$$

	David	*Musician*	*St. Thomas*	*Vitruvian man*
David	0	2.96	3.24	3.54
Musician	2.96	0	3.71	3.79
St. Thomas	3.24	3.71	0	4.26
Vitruvian man	3.54	3.79	4.26	0

Table 1. The 3D-3D mean Euclidian errors (mm) for face shape estimates of Verrocchio's *David*, Leonardo's *Portrait of a musician*, St. Thomas in Verrocchio's *Christ and St. Thomas* and Leonardo's *Vitruvian Man*.

	1	2	3	4	5	6	7	8	9	10	11
David	5.58	6.07	6.16	6.77	5.19	5.94	5.06	5.52	5.71	6.90	7.46
Musician	4.09	5.98	5.47	6.92	4.07	5.83	4.22	5.10	5.72	5.01	5.96
St. Thomas	7.34	8.49	8.03	8.19	5.91	7.62	7.00	5.50	6.49	8.34	7.80
Vitruvian man	6.63	6.10	7.73	6.0	6.54	5.38	6.45	6.46	3.33	8.68	8.91

Table 2. The 3D-3D mean Euclidian errors (mm) between ten distractor faces and face shape estimates of the portraits in Fig. 5.

In turn, the two-dimensional variances are comprised of two components: a) projected three-dimensional variance due to generalization error of the morphable model, and b) two-dimensional pixel error:

$$\begin{bmatrix} \sigma^2_{2D,i,x} \\ \sigma^2_{2D,i,y} \end{bmatrix} = s\mathbf{PR} \begin{bmatrix} \sigma^2_{3D,i,x} \\ \sigma^2_{3D,i,y} \\ \sigma^2_{3D,i,z} \end{bmatrix} + \begin{bmatrix} \eta^2 \\ \eta^2 \end{bmatrix}. \tag{6}$$

The three-dimensional distributions are learned empirically from data and the two-dimensional pixel error is chosen manually. For subjects with more than one image, the error term is summed over all images.

With an estimate of the camera projection parameters to hand, the error term can be globally minimized with respect to the shape parameters by solving a system of linear equations. Similarly, with a shape estimate to hand, the camera projection parameters can be obtained by solving a constrained optimization problem from the 2D-3D point correspondences. Hence, our algorithm proceeds by alternately solving for camera and shape parameters until convergence, commencing from an initialization in which the mean shape is used to estimate camera projection parameters. This is an example of the generalized expectation-maximization or GEM algorithm.[11] This estimation procedure typically converges within ten iterations.

3. EXPERIMENTAL RESULTS

We used our method to construct 3D models of the four candidate portraits shown in Figure 2. The results of this process are shown in Figure 5. The first two columns show the input images used. We extract texture maps from the input images and, in the third column, show this mapped onto the estimated shape, shown in a novel pose. Note that these re-renderings also include frontal illumination showing that the shading (determined by the shape estimate) is in good correspondence with the sampled texture.

We compute mean Euclidian distances between the three-dimensional models from each of the portraits. Because the morphable model is constructed using meshes of known scale, we are able to give Euclidian errors in millimeters. These results are shown in Table 1. The agreement between the David statue and musician portrait is particularly good, compared to the average distances to distractor faces (different subjects of a similar age and primarily male). The results can be seen in Table 2. It is clear that the gross facial shape estimated from the portraits are more similar to each other than to any of the distractor faces.

4. CONCLUSIONS

We have addressed the problem of verifying candidate portraits of Leonardo using three-dimensional face estimation techniques. We view our work as preliminary but promising, showing that the face shape estimated

Figure 5. The three-dimensional digital models inferred from the source artworks. The first (and second) columns show input images and the third column shows a re-rendering of the estimated face shape in a novel pose.

from the portraits is fairly consistent, and are generally more similar to each other than to "distractor" faces. Any firmer claims would of course need to be based on statistical measures of similarities among a large number representative works.

There are many other avenues for potential future work as well. From a technical standpoint, the algorithms could be improved to address the specific requirements of analyzing artwork. For example, one of the strongest

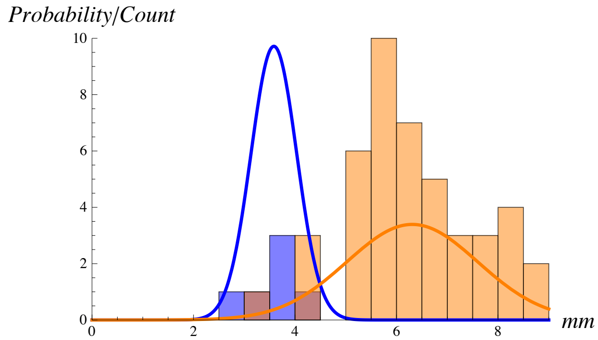

Figure 6. Histograms and best-fit (normalized) Gaussian distributions of inter-face distances within the candidate Leonardo portraits (blue) and between the candidate portraits and ten "distractor" faces (orange), measured in *mm*, as expressed in Tables 1 and 2.

cues, particularly in profile and three-quarter views, is the occluding boundary, which we have not exploited here. In addition, if an estimate could be made of the artist's photometric transformation or style, then it may be possible to exploit texture and shading in the fitting process. Finally, it would be interesting to study the sorts of geometric distortion introduced by artists and due to different facial expression, and to derive a camera model that is capable of capturing this process.

From an art historical perspective, the potential for addressing other questions relating to face analysis is large. As well as comparing and verifying identity, it would be interesting to analyze and modify expression, pose and illumination in portraits. Higher level analysis such as age estimation, and automatic aging of faces might aid in the identification of faces within dated portraits.

REFERENCES

1. G. Vasari, *Lives of the Most Excellent Italian Painters, Sculptors, and Architects, from Cimabue to Our Times*, Torrentino, Florence, 1550.
2. S. Woldhek, *The true face of Leonardo*, Video, 2008.
3. R. Zhang, P. S. Tsai, J. E. Cryer, and M. Shah, "Shape-from-shading: a survey," *IEEE Trans. Pattern Anal. Mach. Intell.* **21**(8), pp. 690–706, 1999.
4. I. Tastl, R. Sablatnig, and W. G. Kropatsch, "Model-based classification of painted portraits," in *Pattern Recognition 1996: Proceedings of the 20th ÖAGM Workshop*, A. Pinz, ed., *OCG Schriftenreihe* **90**, pp. 237–250, Oldenbourg Wien, München, 1996.
5. B. Smith, D. G. Stork, and L. Zhang, "Three-dimensional reconstruction from multiple reflected views within a realist painting: An application to Scott Fraser's *Three way vanitas*," in *SPIE Electronic imaging: 3D Imaging Metrology*, J. A. Beraldin, G. S. Cheok, M. McCarthy, and U. Neuschaefer-Rube, eds., **7239**, pp. 72390U1–10, SPIE/IS&T, Bellingham, WA, 2009.
6. O. Aldrian and W. A. P. Smith, "A linear approach to face shape and texture recovery using a 3D morphable model," in *Proc. BMVC*, 2010.
7. V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *Proc. SIGGRAPH*, pp. 187–194, 1999.
8. O. Aldrian and W. A. P. Smith, "Learning the nature of generalisation errors in a 3D morphable model," in *Proc. ICIP*, 2010.
9. P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter, "A 3D face model for pose and illumination invariant face recognition," in *Proc. IEEE Intl. Conf. on Advanced Video and Signal based Surveillance*, 2009.
10. L. Farkas, *Anthropometry of the Head and Face*, Raven Press, New York, 1994.
11. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*, John Wiley and Sons, New York, NY, Second ed., 2001.